Molecular Cloning and Chromosomal Assignment of the Gene for Human Zn-α_2-glycoprotein[†,‡]

Hisao Ueyama,[*] Deng Han-Xiang, and Iwao Ohkubo

Department of Medical Biochemistry, Shiga University of Medical Science, Seta, Otsu 520-21, Japan

Received June 29, 1993; Revised Manuscript Received September 15, 1993[⊕]

ABSTRACT: Genomic clones containing the human Zn-α_2-glycoprotein gene were isolated. Two of them were for the functional gene overlapped, and the other four were for two different pseudogenes (1 and 2) retaining exon–intron organization. The complete DNA sequence of the functional gene (9.3 kb) and its 5′- and 3′-flanking regions (5.3 and 0.1 kb, respectively) was determined. The gene is composed of four exons; the first exon is for the 5′-untranslated region, the signal sequence, and the first six amino acids; the second, for domain A; the third, for domain B; and the fourth, for domain C and the 3′-untranslated region. The 5′-flanking region contains a TATA box, a CAT box, an octamer sequence, and three possible Sp1-binding sites. Ten and three copies of *Alu* repetitive DNA were identified within the gene and the 5′-flanking region, respectively, and they occupy 30% of the gene. The nucleotide sequences around the exons of pseudogene 1 were also determined; they had high homology (90–91%) with the corresponding region of the functional gene. Southern blot analysis suggested that there are only three genes, including nonfunctional ones, for Zn-α_2-glycoprotein in humans. The gene (ZA2G) was mapped to human chromosome band 7q22.1 by fluorescence *in situ* hybridization.

Zn-α_2-glycoprotein is a glycoprotein found in human blood plasma. It has a molecular weight of 38 000–41 000 (Bürgi & Schmid, 1961) and contains approximately 12% carbohydrates (Ohkubo et al., 1988). Its physiological function is not yet known.

We and others found that the protein is present not only in blood plasma but also in various secretions such as saliva, sweat, and seminal plasma (Poortmans & Schmid, 1968; Frenette et al., 1987; Ohkubo et al., 1990). We prepared a specific antibody against the protein purified from blood plasma and determined the concentration of this protein to be 6.7 mg/dL in blood plasma and 39.6 mg/dL in seminal plasma (Ohkubo et al., 1990). We also examined the distribution of this protein among human tissues immunohistochemically and detected signals in many exocrine glands such as pancreas, prostate, and salivary, mammary, and sweat glands, besides in hepatocytes (Tada et al., 1991). These data suggest that the protein in blood plasma is synthesized in liver and that the one in seminal plasma is not derived from blood but is produced in the prostate gland to be secreted. We therefore purified this protein from human seminal plasma and compared it with the one purified from human blood plasma. The protein in seminal plasma had two unique features: its amino terminus is not blocked, and it contains no carbohydrates (Ohkubo et al., 1990). However, the N-terminal amino acid sequence we determined was the same as that (the second through the 18th amino acids) of Zn–α_2-glycoprotein in blood plasma, suggesting that there is a single gene for Zn-α_2-glycoprotein in humans and that posttranslational modification is different between the organs. On the basis of these data, we isolated its cDNAs from human prostate and liver libraries by using the antibody (Ueyama et al., 1991). These cDNAs of different origin were completely identical.

In this study, we have isolated the human gene encoding this protein, and here we report the complete nucleotide sequence of the gene, including its 5′- and 3′-flanking regions, and its chromosomal localization.

MATERIALS AND METHODS

Materials. DNA restriction endonucleases were purchased from Toyobo (Osaka, Japan), from Takara (Kyoto, Japan), or from Bethesda Research Laboratories Life Technologies Inc. (Gaithersberg, MD). [α-^{32}P]dCTP and [γ-^{32}P]ATP were purchased from Amersham (Amersham, U.K.). The human genomic library (EMBL4) was kindly provided by Dr. M. Takiguchi at the Institute for Medical Genetics of Kumamoto University (Takiguchi et al., 1988). The λ phage DNA was obtained from Toyobo, cut with *Hin*dIII, and used as a size marker.

Screening of Genomic Library. The EMBL4 genomic library was grown in *Escherichia coli* (strain LE392) and screened for gene clones for Zn-α_2-glycoprotein by the plaque hybridization procedure of Benton and Davis (1977). The cDNA coding for human Zn-α_2-glycoprotein (Ueyama et al., 1991) was used as a hybridization probe and was radiolabeled to about 1×10^9 dpm/μg with the random primer labeling kit (Takara) and [α-^{32}P]dCTP.

The hybridization was carried out overnight with radiolabeled probe in 6× SSC[1] containing 5× Denhardt's solution (0.1% each of Ficoll 400, poly(vinylpyrrolidone), and bovine serum albumin), 0.1% SDS, 20 μg/mL each of *E. coli* DNA and RNA, and 50% formamide at 42 °C. The filters were

[†] This work was supported in part by Grants-in-Aid from the Ministry of Education, Science and Culture of Japan (Research Grants 03304028, 02670114, and 05670118 to I.O.) and also by a grant from the Shiga Medical Science Association for International Cooperation (to D.H.-X.).

[‡] The nucleic acid sequence in this paper has been submitted to the DDBJ, EMBL, and GenBank DNA databases under Accession Number D14034.

* To whom correspondence should be addressed.

[⊕] Abstract published in *Advance ACS Abstracts*, November 1, 1993.

[1] Abbreviations: kb, kilobase pairs; SSC, 0.15 M sodium chloride/0.015 M sodium citrate; TBE, 90 mM Tris–borate (pH 8.3)/2 mM EDTA; bp, base pairs; BrdU, bromodeoxyuridine; FITC, fluorescein isothiocyanate; MHC, major histocompatibility complex; HLA, human leukocyte antigen.

0006-2960/93/0432-12968$04.00/0

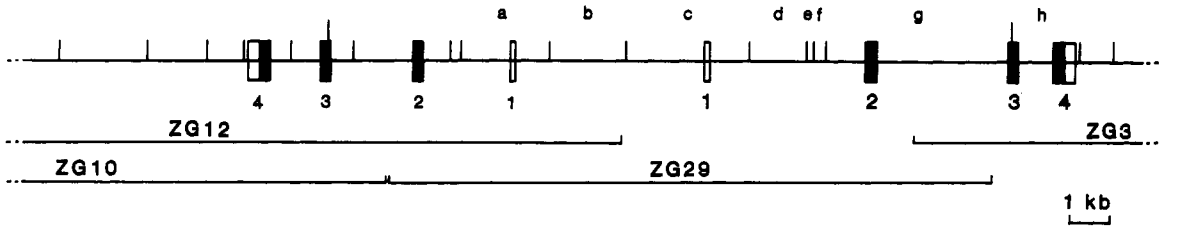


FIGURE 1: Structure of the human Zn-α_2-glycoprotein gene. Boxes indicate exons, which are numbered 1–4; the closed portions are for mature protein-coding regions and the open portions are for the signal sequence and untranslated regions. The right-hand set of exons is the functional gene, and the left-hand set is a pseudogene (pseudogene 1). ZG3, ZG10, ZG12, and ZG29 are phage clones for the gene, and the range of coverage of their inserts is shown. Vertical bars indicate *Eco*RI sites. The *Eco*RI fragments are designated a–h and correspond to those in Figure 2.

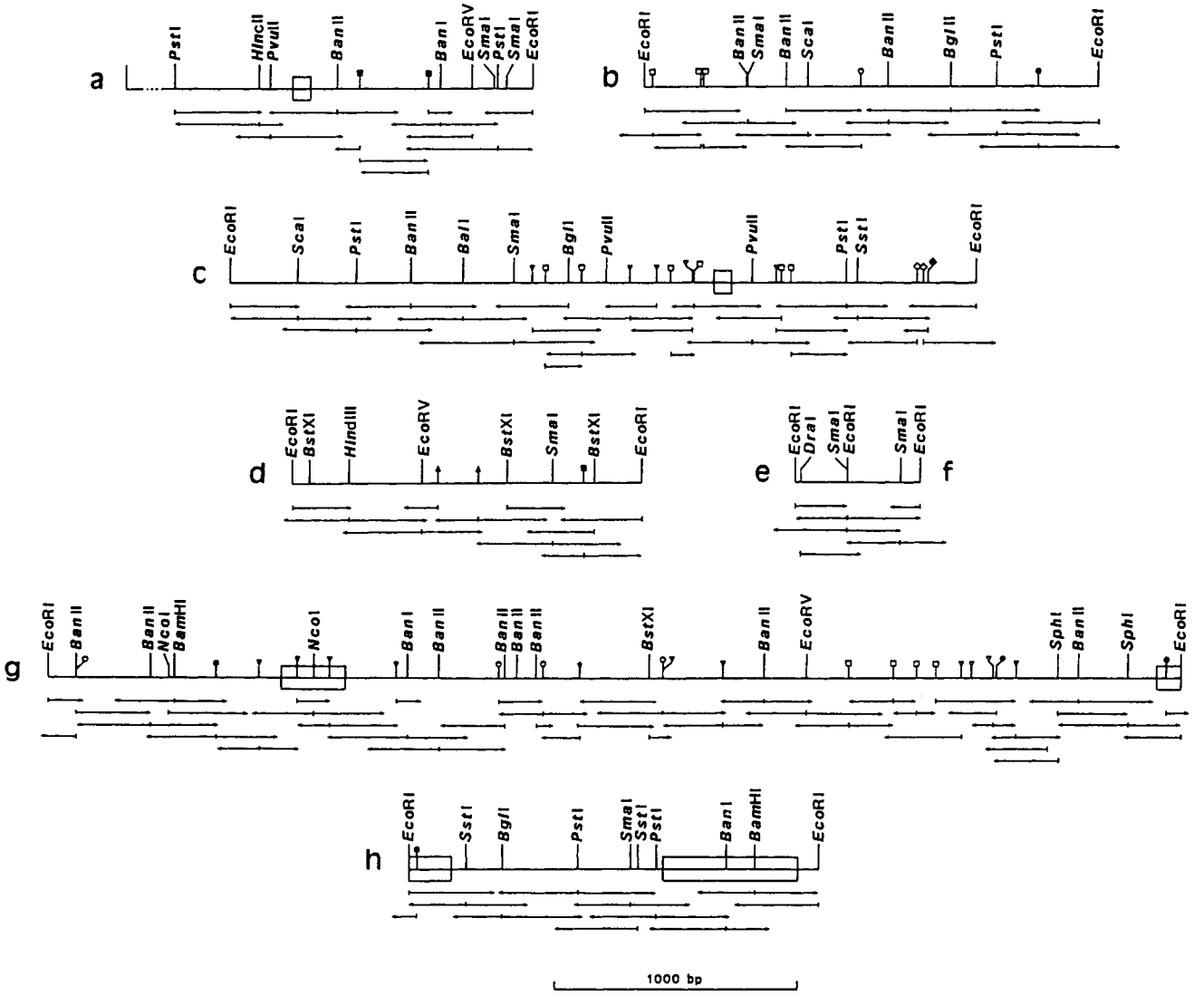


FIGURE 2: Detailed map and sequencing strategy for the human Zn-α_2-glycoprotein gene. The boxes indicate exons. Restriction sites for six base-recognizing enzymes are all shown within each of subclones a–h, which correspond to those in Figure 1. Restriction sites for four base-recognizing enzymes are shown only when they were used for subcloning: ■, *Alu*I; □, *Hin*fI; ○, *Msp*I; ●, *Taq*I; ▼, *Dde*I; ▽, *Hae*III; ◆, *Rsa*I; ◇, *Acc*I; △, *Ava*I; ▲, *Hha*I; *, *Mbo*II. The extent and the direction of sequencing by the dideoxy chain termination method are shown by arrows; the vertical bars associated with the arrows indicate the starting sites for DNA sequencing.

washed twice with 6× SSC at 42 °C, twice with 2× SSC containing 0.1% SDS at 65 °C, and four times with 0.1× SSC containing 0.1% SDS at 65 °C sequentially for 15 min each. Phage clones with positive signals were subjected to further screening for purification. The phage DNA was isolated from liquid culture and cut with *Eco*RI. The digested DNA was run on an agarose gel (0.7%) and blotted onto a nirocellulose filter, which was then hybridized with three different probes: the 5′-region (*Hin*fI, up to the 52nd amino acid), the middle region (*Hin*fI–*Eco*RI, the 53rd through the 126th amino acids), and the 3′-region of the Zn-α_2-glycoprotein cDNA (*Eco*RI–*Bam*HI, from the 127th amino acid to the C-terminus and a part of the 3′-untranslated region) (termed N, M, and C probes, respectively). According to the electrophoretic pattern, the phages were classified into six groups, and one representative of each was subjected to further analysis.

DNA Sequence Analysis. The DNA sequence was determined by the dideoxy sequencing method after DNA fragments were subcloned into a plasmid vector, pUC119, with the Sequenase sequencing kit purchased from United States

```
-5301  CTGCAGTGGT GTCTGAGTGT GGGGTGAGGG CACGGGTTTC CTTCTAGCAC CATCCCTCCA GCCAACCTGC CTTCCCACTT TCCCCTTGGG GGCTGTTCTC
-5201  TCTTTTCCTC AGCGTCAGGC AGGTGTACCC AGCCATCGGA AGGAGGGAAG GAACAGGAGC AGAAGCAGCT CCAGGCCTGT GGTTCCTGCT CCGTCCTGGC
-5101  TCTCCACCCC AAGCAGGTCG CTTGGTCTTT GGAGTCTTTG TTTTCACTGT AAACTAAGGG GTTTTAGATT CAGATTCCTG AAGGTCTTAT TTTGGGTGAG
-5001  GGCAGAGGCT CCCTGAATGC TCCTGGAAGT GTGTGAGCTG AGGGAGTTGA CAGCCTTGGA CACCTGTTCC TGCCATATGC CAGGGAATCC CAGCTGCATC
-4901  CAGCCCTTCT CCCAGCCACA TGTCCTGTTC CTCACCTCCT TCCAGTCCTC TCCAGCACCC ATCCCTTGCT TTCCCCACTC ACCATCTCGG GTCTCCTGGA pseudo
-4801  GGACAGCAGG ACCCAGAAGC AGCAGCAGAG ACAGCAGGAC AGACACCATT CTTACCATTG TGTCTGCTTG GAAGGTTCTG GGCAGGAGGC ACAGGTATTA Exon 1
-4701  TCCTGGGTCC TGGGGCCATG CCCAGGGGAG GTGAATCTAC AGGCCAATGG GAGTACCAGG CCGGGCTCCT CCCCTGACAG TAAGAAAGAA GTATGCACAA
-4601  GCCAGGCTGG GCAGAGGTTG TGATCAGTGA GTCTACACAC AGGATATGCA AATTCAGAGC TGAGACAGTT GATCTCTGGA CTGAGGATTT GTGTGTGCCT
-4501  GGGGCAGAAG GATGCCAGTG TGTGGGAGGT GTGTGTGATG GGTGGCAGGT GCCTGGGATG TGTCGCCCTT CCAGTCTATT ATGACACCCT CTGGGGTCAC
-4401  TTGGGTGTTG TCAGACCACA CTTCATCTCT GCCTTCTGGA AAAACTCTCA ACATGTCCAA GACACGGACA TGGGTAGACC CCCGCTGGGG TGGGACCCTC
-4301  ACTCTCCCAC ATCCGATCCT GCCAGAGCTG TCTGGAGGAC ATTCCCATTC CCATCTCTCC CTCTTTCCTT TAGGTGCCTG TGGAATCCCA GCCTTCCATG
-4201  ATGGAAGCAG AAATAGCCGC ATCCTCCTTT TGCAGCCTCC CTGGCAATAG AACGAGCAGG TGTGCACAGG GTTTATGCTC AGCAAGTCGA AGCTACTGCC
-4101  TTGGATATCG GGTCAAGTGT TAGGGAAACA AAGAGGAGGG CACAGTGGAC AATGTTTTCT GGTGGTCTTG GCAGGGGTTG CAACTGCCAG CTCCCGGGGC
-4001  AGTGGGGCTG CAGTAGCTGC GGGGGCGTCT TGAGCTTCAG TCCCCCGGGA TGTTGTAGAT CCCGTGCTGT GGCCTCTGCT GCCCCCTGCT GGCAGCTTCT
-3901  GTATATTTAC TGGGCCGGTT GCAGGTGTGA TTTTGACTGT TGGGCTCCTC TGGAATTCTG AGCCTGGGTT CTCCATCCTT TCTGTGGATT CTGAGAGCTA
-3801  ACCAATAAAT ACGTTTATTT TCTGCTTAAG GTCAGAACTG CATTACGTTT CCAATAACAA ATACTCCTGG CGGATACAGA GATTGGTACA AGGATTGTTT
-3701  CCAGGCAGTA TTTACTCAGG GAAATGGAGG GCTGGGAAGC TGAGGTTGGT TACTGGGCCA GATTGGGGCT GCGCAGTGAA AACACAAGAG TCTTGGGGGA
-3601  TTCTGGCAGG CCTGGTGTGC AGTGATAACT TCAACAGTTG GCTGAGGTAT CTGTTATGGT AGATTGGTGT AGGGACCAGC CCCACAGGGT CGGAGGGTTT
-3501  TTCTCCCCGT GTGCAGAGAC GAGAGAGTGT AGAAATAAAG ACACAAGACA AAGAGATAAA AGAAAAGACA GCTGGGCCCC GGGGACCACT ACCACCAAGA
-3401  CGTGGAGACC GGTAGTGGCC CCGAATGTCT GGCTGCGCTG TTATTTATTG GATACAAAGC AAAAGGGGCA GGGTAAAGAG TGTGAGTCAT CTCCAATGAT
-3301  AGGTAAGGTC ACGTGGGTCA TGTGTCCACT GGACAGGGGG CCCTTCCCTG CCTGGCAGCC AAGGCAGAGA GAAAGAGGAA GAGAGAGAGA CAGCTTATGC
-3201  CATTATTTCT GCTTATCAGA GACTTAGTAC TTTCACTAAT TTGCTATTGT TATCTAAAAG GCAGAGCCAG GTGTACAGGA TGGAACATGA AGGCGGACTA
-3101  GGAGCGTGAC CACTGAAGCA CAGCATCACA GGGAGACGGT TAGGCCTCTG GATAACTGCG GGCGGGCCTG ACTGATGTCA GGCCCTCCAC AAGAGGTGGA
-3001  GGAGTAGAGT CTTCTCTAAA CTCCCCCAAG GAAAGGGAGA CTCCCTTTCC CGGTCAGCTA AGTAGCGGGT GTTTTCCTTG ACACTGAGGC TACCGCTAGA
-2901  CCACCGTCCG CTCGGCAAGG GGCGTCTTCC CAGAACACTG GCGTTACTGC TAGACCAAGG AGCCCTTCTG GTGGCCCTGT CTGGGCATAA CAGAAGGCTC
-2801  GCACTCTTGT CTTCTGGTCA CTTCTCACTA TGTCCCCTCA GCTCCTATCT CTGTATGGCC TGGTTTTTCT TAGGTTATGA TTATGCAGCG AAGATTATTA
-2701  TAATATTGGA ATAAATTATA ATATTGGAGG ATTATTATAA TATTATAATA TTGGAATAAA TAAATTTGGA AAAAACTAAT GATTAATGAT ATTGATATAT
-2601  AATCATATCT ATGATCTAGA TCTAGTATAA CTCTTGTTGT TTTATATATT TTATTATACT GGAACAACTC ATGCCCTCGG TCTCTTGGCT TGGCACCTGG
-2501  GATGGCTTGC CGCCCACATC TCCCCCCTTT TTATTAACTA GGATCGCCAT CGCCATCATT GCTTGTCGTT GACTTCGGAT TTGTTTTCGG ACTCCTTGGA
-2401  GGCATCTGCA GGCTAAAAGG AGACAACGTA AGCATACCAA TATTAATAAT GCCAGTGACA ACAATGATCC TCCAAGGGGT TTGATCTATT TAAAGGGATT
-2301  AAGATCAGAT AATTGTTTAG TTATGCCTTC AAAAATGTCT GAGCCAGGAA CAGTGGATAA ATGAGCTTGT GAATCCTCGA AAATTTGCTC TTTAAGTTTT
-2201  GAAATATCCA AGGTTAAGTT ATCATCCCAG GCTTTTAAAT GTCTTGAGAC ATTTTCCCAG CTATGTTGAT ATTTATTATA AGCATAAGGC ATTATGCAAT
-2101  AATCAGAAGT ATTCCAATCA CTCTGTAATT GCATACGGTG TTCCAAATTC ATATCTCCCA GCCAGATTAC ACTTTGGCGG AGATCATTAA TTTGATTAGC
-2001  TAATTTTTGA TCAATTTGAG CCTGAGAATT CCAAAGTCTG GAGGAGTTTT TCTGCCATGC TTCAATTCAA CATATTGAAC GGTTTGAACA GAATTGTGGA
-1901  TGGCAACTCC AGCAGTTGCT GCTGTTGCAG TAACCGCAAT TAGTCCTGCA ATGATGGCAA TAAGAGTAAA AATAAATCTC TTTGTTCTTT TGCGGATACC
-1801  TTTAAGAACT TCATTGACTA TGTGAATAGA GGGGGAAGAC TCCCATGGAC GATGTAAAGA AACTGGTATC CATACCCCCT CCCTAGCCCT TACCAAGAGA
-1701  GTACTTGTAG TGGGATTAAA AGTAGCATCA ATGCACGTGA ACAGCTTACA GTTATCACAT TCTATAGTTT GTGTATTGGG AATGATAATT ATATTTCCAA
-1601  CCAACAGCAT GTAAGGGGGT TTGACATAGC TCCTGATGGG TATCACCCGT TCAGATATGA GGGTGATGTT GAATGTGGGT GTTTTGGTAT TAGTGTGAAG
-1501  GAGTTGATAG GTAGTGTTCC ATATCCTTAT TCCTGTCATA GCTGCAGCTA ATTTTCATAG TTGAGGATGT TCTGGGGTAA CAAAGGGATG AATCATTTTT
-1401  GGTCTAGGAG GAGTAATGCC TGCATCCATC CATTTAAATA GGTAAGGGGA CACCCATTGT CTCAGCCTGT ATGACTGCCA TCCATCCTCT ACATAATCTA
-1301  ACAAATAATT AAACTCTGAG CATGAGGTAT TTTTGCTGGA GCAATCTTGC CAATAATGCC CTTTTGGAGC CCAATCAATA ACAATACCTG CGGCTAGACT
-1201  TTGGAGCACA ACGGCCTTAG GAGCATTACA ATTATTCCAT AAAATTGAAG TCACTATAAA AGGTCCCTTT GTAGGTTTCT TTAACCAAAA GAGAGGCTGG Alu 1
-1101  GCACAGTTGC TCACACTTGT AAACCCAGCA CTTTGGGAGA CCGAGGCGGG CAGATCACGA GGTCAGGAGT TCAAAACCAG CTTGGCCAAC ATGGTGAAAC
-1001  CCCTGCCTCA GCTTCCTGAG TGCTGGGATT ATAGGCATGT GCCACCACGC GCAGCTAATT CTTGTATTTT TTCAGTAGAG ACGAGGTTTC ATCATGTTGC Alu 2
 -901  CCAGGCTGGT CTCGAGTACC TGACCTCAAC CTGAGGTGAT CCAACCACCT CAGCCTCCCA AAGTGCTGGG ATTACAAGCA TGAGTCACCG CACCCGGGCC
 -801  CCAGTCACTT TAGAATAGCA TGTTGCTTTG TATTCGGAGG GTCTCTCTGC AAATAGCCCA TCAACACTGA GCGTGCCTGG AAAGACCTGG TTTTCAAATA
 -701  ACTGGCTTCG TCTGTGTAAA ACGAGTCTTG TTGTATGCAT TAAAAATTAT CTTGGCTGGG CGTGGTGGCT CACGCCTTTA ATCCCAGCAT TTTGGGAGGC Alu 3
 -601  TCGTTCTGTT GCCAGGGAGG CTGCAAAAGG AGGATGCGGC TATTTCTGCT TCCATCATGG AAGGCTGGGA TTCCACAGGC ACAGAAAGGA AGGAGGGAGA
 -501  GATGGGAATG TGACTGTCCT CCAGACACAG CCTCTGGCAG GATCGGATGT GGGAGAGTGA GGGTCCCACC CCAGCTGGGG TCTACCCAGG TCCATGTCTT
 -401  GGACATGTTG AGAGTTTTTC TGGAAGGCAG GGATACAGTG TGGTCCAAAA ACACACAAAT GCCCCTACTG GCCCAGGGGT TGTCACAATA GACTGGAAGG
 -301  GTGACACATC CCAGGCGCTT GCCACCCATC ACACGCACCT CCTACCCACT GGCATCCTTC CACCCCAGGC ACACACAAAG CCTCAGTCCA GAGATCAACT
 -201  CTGGACTCAG CTCTGAATTT GCATATCCTG TGTGTAGATT CATTCTTCAT AACCTCTGCC CAGCCTAGCT TGTGTATCAT TTTTTTTTCT CTATTAGGGG
                   octamer                                Sp1
 -101  AGGAGCCCGT CCTGGCACTC CCATTGGCCT GTAGATTCAC CTCCCCTGGG CAGGGCCCCA GGACCCAGGA TAATATCTGT GCCTCCTGCC CAGAACCCTC
       Sp1                    CAT                     Sp1                   TATA

       +1         MetValAr gMetValPro ValLeuleuS erLeuLeuLe uLeuLeuGly ProAlaValP roGlnGluAsn GlnAspG
   -1  CAAGCAGACA CAATGGTAAG AATGGTGCCT GTCCTGCTGT CTCTGCTGCT GCTTCTGGGT CCTGCTGTCC CCCAGGAGAA CCAAGATGGT GAGTGGGGAA Exon 1
  100  AGCAAGGGAT GGGTGCTGGA GAGGACTGGA AGGAGGTGAG GAACAGGACA TGTGGCTGGG AGACAGGCTG GATGCAGCTG GGATACCCTG GCATACGGCA
  200  GGAATGGGTG CCCAAGGCTG TCAACTCCCT CAGCTCACAC ACTTCCAGGA GCATTCAGGG AGCCTCTGCG CTGGCCCGAA ATAAGACCTT CAGGAATCTG
  300  AATCTAAAAC CCCTAGTTTA CAGTGAAAAC AAAGACTCCA AAGACCAAGC GACCTGCTTG GGGTAGACAG TCAGGACGGA GTAGGAACCA TATGCCTGGA
  400  GCTGCTTCTG CTCCTGTTCC TTCCCTCCTT CCGATGGCTG GGTACACCTG CCTGACGCTG AGGAAAAGAG AGAGCAGCCC CAAGGGGAAA GTGGGAAGGC
  500  AGGTTGGCTG GAGGGATGGT GCTAGAAGGA AACCCGTGCC CAAATCCCAC ACTCAGACAC CACTGCAGTG GGTCTGGAAG GCGAGTGGCT GGAAGAGAAG
  600  AGAGTGGGAG CTCCGGGAGA TCAAGAGTCA CTCCTAGGAT AAGGGAAGGA GGCTGTTTGT GGCATGAGAA TGTGCAGGAT AAAGACATGG AAGCGAATGG
  700  CTTCTCAGTT GTGTGAGTTT AAAATTCATG ACATTTACAA ATTGTCAGAA AAGGTGTTAT ATGTTTGTTA TATAACAATC ACTTTGGAAT GTTAATCTGA
  800  TTCTGTGCCA AAATCTGAAT TACTCAGGGT TCTCCAGAGA AACAGAACTA ATAGGTGGTA CACATATACA TATATATGTA CGTACACATA CATACATACA
  900  CTGTATACAC ATGGATACAC ACACACATAG GAAGAGATTT ACATATATGT ATACAAAAGA GAGAGAGAGT AGAGATTTAT TTTAAGAAAT TGACTCACAC
 1000  TATTGGGAGG AGTAACAAGT CCTAAATCTT CAGAGCCGGC CAGCAGGCTG GAGACCCAGG GAAGAGTTGA TGTCTTAGTC TTGATTCCAA GGGCAGACTG
 1100  TAGGCAGAAT TCTTTCCTCT TTAGGGGACA TCTGAGGCTT TTTCTCTTAA GGCCTTCAAC TGATTGGATG AAGCCCACCA CTATGGAGAG TAATCCACTT
 1200  TACTCAAGGT CTACTGATTT TTTTGTAAAT TAAAAAAAAA ACTGTGGGTG CATAGTATGT GTATATATTT ATGGGGTACA TGAGAGGTTT TGATTCAGGC
 1300  ATGCAATGTG AAATAATCAC ATCATCAAAA ATGAGGTATC CATCCCTTCA AGCTTTTATC GTTTGTGTTA CAGACAATCC AATTATACTT TTTTGGTTAT
 1400  TTTAGTTTTT AAAAGTATTT GATTATTTAT TTATTTATTT ATTTTTGAGA CAGAGTCTCA CTCTGTCACC CAGGCAGGAG TGCAGTGGCA TGATCTCGGC
 1500  TCACTGCAAC CTCCGCCTCC CAGGTTCAAG CAATTTTCCT GCCTCAGTCT CCTGAGTAGC TAGGACTACA GGCACCTGCC ACCACACCTG GCTAATTTTT Alu 4
 1600  TTGTATTTTT AGTAGAGACG GGTTTCATCA TGTTGGCCAG GCTAGTCTTG ATATCCTGAC CTCGTGATCT GCCCGCCTTG GTCTCCCAAA GTGCCGGGAT
 1700  TACAGGTGTC AGCAACTGCG CCTGGCCTCT CTTTTGGTTA TTTAAAAGTG TACAATTAAA TTATGATTAT TATTATTATT TTTGAGATGG ATTCTTGTTC
 1800  TGTCACCCAG GCTGGAGTGC AGTGGCGTGA TCTTGGCTTA CTGCAAACCT CCGCCTGTTG GGTTCAAGCA ATTATCTTGC CTCGGGTGTA CACTGCCACA Alu 5
 1900  CACGGCTAAC TTATGTATTT TTAATAGAGA TAGGGTTTCA CCATGTTGGC TAGACTGGTC TTGACCTCTT GACCTCAAGT GATCCACTCA CTTCAGCCTC
 2000  CCAGAGTGCT GGAATTACAG GCACGAGCCA CCACACCTGG CCCCAGTTAA ATTATTATTG ACTATAGTCA CCCTGTTGTG CTATCAAATA GTAGGTCTTA
 2100  TTCATTCTTC TTTTTTTTTT TTTTTTGTGA CAGAGTTGCC CAGGCTGGAA TGCAGTGGTG CAATCTTGGC TCACTGCAAC CTCTGCCTCC CGGGCTTAAG
 2200  CGATTCTCCT GCCTCAGCCT TCTGAGTCGC TGGGACTACA GGTGTGTGCC ACCACGCCCG GCTAATTTAT GTATTTTTAG TAGAGATGGG GTTTCACCAT Alu 6
 2300  GTTGGCCAGG CTGGTTTCGA ACTCCTGACC TCAAGTGACC CACCTGCCTC AGCTTCCCAA AGTGTTGGAA TTACAGGCAT GAGCCACCAC ACCTGGCCCC
 2400  AGTTAAATTA TTATTCACTG GAGTCACTTT GTTGTGTGCT ATCAAATAGT TTTCTAACTA TTTTTTTTGT ACCCATTAAC CACCCTCCCA ATTTCCCCCC
 2500  AACCCTGCCA CTACCCTTCC CAGCCTTTGG TAACCATCCT TCTACTCTCT ATGTCCATGA ATTCAATTGT AGGGTCTACT GATTTAAAGG CTAATCACAT
 2600  TTAGACACTC AGGAGCAAGA ATAATTTTAG TAATTGAACT AGGATTCTGC CATATGACCT CCAACATCAT TAGCACCTGT GTAAATTGTA TCATAAAATA
 2700  ATTATGGAAC TATTATGGAA ATGTCCCTCT CTCCCAGATC CCACCTTGTA CCAAAATGCA AGGTACAACC CCGGGAATTC TGAGCTCCAT CCTAGTCTTA
 2800  CCCTGTGCTA ATTCAGTCTG GGTCATTTCT TGAATTTTCT GGTAAATTCT CCTTTCTACC CTTTCTAACT ATATGTATTT GTCAGGTTAA GCTAGAAGTG
 2900  TTAATTTTTT TTTTTTTTGA GATGGAGCCT TGCTTTGTCA CCTAGGCTGA AGTGCAGTGG CATGATCTCA GCTCACTGCA AGCTCCGCCT CCCGGGTTCA
 3000  TGCCATTCTC CTGCCTCAGC CTCCTGAGTA GCTGGGACTA CAGGCACCCG CCACCATGCT TGGCTAATTT TGTGAATTCT TAGTAGAGAC GGGGTTTCAC Alu 7
 3100  CATGTTAGCC AGGATGGTCT CGATCTCCTG ACCTCGTGAT CCACCCGCCT CGGCCCCCTA AAGTGCTGGG ATTACAGGCG TGAGCCACTG AGCCCGGACG
 3200  AAATGTTAAT TTGTTTTTTT TGAGACGGAG TCTCACTCTG TCATCCAAGC TGGAGTGCAG TGGCATGATC TTGGCTTGTT GCAACCTCTG CCTCTCTGGT
 3300  TCAAGTGATT TTCCTGCCTC AGCCTCCAGC ATGACTGGGA TTACAGGCCC GCACCACCAT GCCCAGCTAA TTTTTGTATT TTTTAATAGA GATGGGGTTT Alu 8
 3400  CACCATGTTG GCCAGGCTGG TCTTCAACTC CTGATCTCAA GTAATCTGCC TGCCTTGGCC TCCCAAAGTC CTGGGATTAC AGGCATGAGC CACGGAGCCC
 3500  AGCCTAGAAA TGTTAATTTC TAACGCATGT CAGATTCCAT GCACACTGGG CAAGGTTCCA TTCCTCCATG GGGTGACTCA GGGATCCAGG CCAATTGCAT
 3600  ATTGAGACTC TTTCATATTA TCCTGTGGCC TTCAAAGTCG TCACCTCTAG GGATGAGAAA CAAAAGGGAC AAGCCAGCTG GTAGGGTCTT GGACAAGAAG
 3700  AAAGACATCA CTTCTGCTCA CATTCTCTTT TGACAAAACT CAGTCACATG GTCCCAATAT ATCTTCGAGG TGGCTGAGTA ATGTTATCTT CCTATGTGTC
```

```
3800  AAGCAGAGGA AATAATGTAG TGAAGACACA GGATGGTCTC TGAAATATCA TCTCAGGCAT GAAAGTAGAG CATATTCACT TGAGTGAGCC TCCAGTGGTG
3900  TGAAGTTGAT GGCAGGAGAA AGAGCTGGGG AAGAAAAGGC CAGTGGCAGG TCTCCCCTCC TAGCCCTATG CAGCCCCACA GTGGGACCCT TGCATGGACC

                                  ly ArgTyrSerL euThrTyrIl eTyrThrGly LeuSerLysH isValGluAs pValProAla PheGlnAlaL
4000  TCAACCATCA GAATCTTTTC TTTTGCAGGT CGTTACTCTC TGACCTATAT CTACACTGGG CTGTCCAAGC ATGTTGAAGA CGTCCCCGCG TTTCAGGCCC

      euGlySerLe uAsnAspLeu GlnPhePheA rgTyrAsnSe rLysAspArg LysSerGlnP roMetGlyLe uTrpArgGln ValGluGlyM etGluAspTr
4100  TTGGCTCACT CAATGACCTC CAGTTCTTTA GATACAACAG TAAAGACAGG AAGTCTCAGC CCATGGGACT CTGGAGACAG GTGGAAGGAA TGGAGGATTG Exon 2

      pLysGlnAsp SerGlnLeuG lnLysAlaAr gGluAspIle PheMetGluT hrLeuLysAs pIleValGlu TyrTyrAsnA spSerAsnG
4200  GAAGCAGGAC AGCCAACTTC AGAAGGCCAG GGAGGACATC TTTATGGAGA CCCTGAAAGA CATCGTGGAG TATTACAACG ACAGTAACGG TCAGTGAATA
4300  ACAGACCACA GGGGTGGAAG GTCTAACCCA AGAGGCAGCC CCCCCAGTGT GAGTGGCAAG GGATCAGCAG GATGGAAATA GTCCCAATCC CAGGGGAAGA
4400  ACAGGAGACA CAGCAGAAAC ACAGACATGT CCACATCCCA CCCACCCCAC AGCACAGGTG CTCCCCGCTT CCCCATCAAT TGCCCCATCC TCATCCCAGG
4500  CCTCAGGTCA CACAGGAAGT GATGGCAGAG TCACTTCCTA TCCAGGCACC TATGACCTCT CACCTCCACA CCCCACCCAT CGGAGGCTGA TACCCCCGTG
4600  AGAAGGCATC AGACTCACCC CTGTCCAGGG AGGTTGCCTG GAGAGTGAGC CACTCTCAAA GTCACTCAGA CCTGGGCTCA CCTGGTGGTT CTGCCAGTCC
4700  TAGCTGTTGA CAGTGAAACG TTCCCAAAAT ATCTGGTTGA AATCTGCAAA CATTGGAGCA CTGAGACCTA CCTCCAAACA AGTCTGTAAT ATTTAACTAT
4800  GTCTGTTCTA TAAGGATGTC ACAGTCTGTC CTGATCTCCC TTGCAGCTCC ATCACCTAGC ACAGGGTACA GCCAATATTG GCTCAATTGA AATTTGTGGA
4900  ATCCACAGAG AAAAGCACCC GGCACACACC GTAGCCCATG CTGGGGGCTC AGGAAGTGCT GGATTCAAAA CTGTGGGCTG TTAGAGTTCC CTTGGAGCCC
5000  TAAAGTTCCT CCTTACCATA CGATGCAGAC CCAGGAAGGG CCACCTGCGC TATGGTCAGA GGAGCTGGTG GCAGAGCCCG TGCAGAGATG GTCCCTGTGC
5100  CCCCGGCCCA GTGCTCTTTC TCCTAAACCA CACTGCCAGC CCCAAGGCAG CCAACCTCAG GTCTGGTGAA CTGCTGGTGT TAAATTATCA TAGAGTGGGT
5200  GTCAAAAGAT GGGCTACTAA GTACAAAAAT GCCCAAGGTG CTACATGGGA TCTGAAGATT TTCAAAAGGA GGCAAGAAAG AGATAGGCAG ATGTTTCAAG
5300  GATGTGGGGT GGGGGAGGTC TTGGTAAGGA AAATGGCCCA GGCTGTGTGT CAGCAATAGG AGAGGAGGGG GCACAGGTGA TCAGAAAAGA CACTGGGGGA
5400  AGCATTGATG GACAGGAATA GAAATGGCAA AGTGGATAAT TAAGAGGAAG GAGGATGAGG AGATGAACAC AGGGTATTAG AAAATAATAG AAGGCAGGGC
5500  TTGGTGGCTC ACTCTTGTAA TCCCAGCACT TTGGGAGGCT GAGGCAGGCA GATCACCTAA GGTCAGGAGT TCGAGACCAG CCCGGCCAAC ATGGTGAAAC
5600  CCTGTCTCTA CTAATAATAC AAAAATAGCC TGGCATGGTG GCACACGTCT GTGGTCCCAG CTACTCAGGA GGCTGAGGCA GGAGAATTGC TTGAACCCAG Alu 9
5700  GAGGCAGAGG TTACAGTGAG CCAAAATCCT ACCATTGCAC TACAGCCTGA GTGACAAGAG TGAAACGTTG TCTAAAAACA AAAAACAAAA AAAGGAAATA
5800  ATAGTAGCTG ACATTTACTG AGCACTTACT TTGTGCCAGG CCCATCTATG AGCATATATA ATGCTCAGAA TAGCCCCCTA AAACAGTGCT CTTGGCATTG
5900  CCATTTCAGA GGTGAGGAAA TAGAGGCACA GGGAGTTGAG TGGCTCCAGT TCAGGCAACA CACCAGGTGG GGGTGGGGGG CTGGGGAGAG ACCTGGGACG
6000  TGAGCCCAGA CAGCTTGAGA GCTTTCAGAG TCTATGCCAA CAGCACCAAC CAGTGCTGGG TAAACACCTG CTTTTATCAT CAGAACAAAG AGGCTGTGTC
6100  CCCTGCCCTA TGAGGTCCAT TTCTGAGAGT TGTGGCTAAT GGGCAAGAAG GTTGGGGCTT TAGAGATTTG GGATAAAGAT ATCAAACACC AGAAAAGTAG
6200  AAAGAAGTGA TCAGATTAGG GTTACTTAGG TGATGATATG AACTCTTCCT AGAACTGAGA GAAAAAGAGA GCCTTCCTTT ACTCATATGA AATCACAAAT
6300  AATTTCTATC CAATTTGGAA GTACACTTTG GTGTAGTTGT GACAGCTTCC TCAGGACTCA GCATAAATTC AAACAAATAA TTGTCCTTAG AAGAGATGCT
6400  ATAGAAGAGA TAGAAATATA TTCATATTCT GTAGCTTTTT TTTTTTTGAG ATGGAGTTTT GCTCTTGTCA CCCAAGCTGG AGTGCAGTGA TGCAATCTCA
6500  GCTCACTGCA AACTTTGCCT CCTGGGTTCA AGGGATTCTC CTGCCTCAGC CTCCCGATAA CTGGGACTAC AGGCTACAGG CATGTGTCAC TACTCCTGGT Alu 10
6600  TAATTTTTTT TTTTTTTTTT AAGACTGAGT CTTGCTCTGT CTTTCAGGCT GATGTACAAT GGCTCCATCT CGGCTCACTA CAACTTCTGT CCCCCAGGTT
6700  CAAGCGATTC TCCTGCCTCA GCCTCATGAG TAGCTGGGAT TACAGGCATG TGCCAGCACA CCCAGCAAAT TTTTGTATTT TTAGTAGAGA TGAGGTCTTA Alu 11
6800  CCATGTTGGC CAGGCTGGTC TCAAACTCCT GACCTCAGGT GATCCTTTGG CCTCAGCCTC CCTAACTGCT GGGATTACAG GCATGAGCCA CTGCGTCCAG
6900  CCTAATTTTA TATTTTTGGT AGAGATGGGG TTTCACCATA TTGGCCAGGC TGGTCTCGAA CTCATGACCT AAGGTGATCC ATCCTCCTCA GCCTCTCAAA Alu 12
7000  GTGCTGGGAT TACAAGTGTG AGCCACTGGG CCTGGTGCTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTGAGATA GGGTCTCACT CTGTCACCCA
7100  GGCTGAAATG CAGTAGTGTG ATTTTGGCTC ATTGCAGCCT TGACTTCCCA GGCTGAAGTG ATCCTCCCAC CTCAGCCTCC TGAGTAGCTG GGGCTACAGG Alu 13
7200  CATGCACCAC CATGCTGCGC TAATTTTTAT ATTTTTTGTA GTGGTGGGAT TTTGCCATAT CACCCTGGCT GGTCTGGAAC CCCTGGGCTC AAGCGATCCA
7300  CTCGCTTCAG CTTCTCAAAG TGCTGGGATT ACAGGCATGA GCCACAGCGC CCAGGCTGTA GCTCTCTTAA GGAGGAACAT ATCTCATCTG AGACAAACCT
7400  GAAATGCCAA ACCAAACTGA GTTAGCCCCT CTCTGTCTGT TGTATATATT GGAGTAATAA CCTATTTGTC TTGATAAAGG GATTGCATGC TTGAATTGCA
7500  AAAACCTTTA TTTCTTTTGG GTTGCCCAAT GTGCAAGACT AAGAGTTATT TTGATAAATT TCTCACCAGG CTGACTGTCT CTCTGTGGGG TCGGGGGAGT

           lySe rHisValLeu GlnGlyArgP heGlyCysGl uIleGluAsn AsnArgSerS erGlyAlaPh eTrpLysTyr TyrTyrAspG lyLysAspTy
7600  TTTCAGGGTC TCACGTATTG CAGGGAAGGT TTGGTTGTGA GATCGAGAAT AACAGAAGCA GCGGAGCATT CTGGAAATAT TACTATGATG GAAAGGACTA

      rIleGluPhe AsnLysGluI leProAlaTr pValProPhe AspProAlaA laGlnIleTh rLysGlnLys TrpGluAlaG luProValTy rValGlnArg
7700  CATTGAATTC AACAAAGAAA TCCCAGCCTG GGTCCCCTTC GACCCAGCAG CCCAGATAAC CAAGCAGAAG TGGGAGGCAG AACCAGTCTA CGTGCAGCGG Exon 3

      AlaLysAlaT yrLeuGluGl uGluCysPro AlaThrLeuA rgLysTyrLe uLysTyrSer LysAsnIleL euAspArgGl nA
7800  GCCAAGGCTT ACCTGGAGGA GGAGTGCCCT GCGACTCTGC GGAAATACCT GAAATACAGC AAAAATATCC TGGACCGGCA AGGTACTCAC TGCTTCCTGC
7900  TCCCCAGTAC TGAGCCCAGA ATAAAAGACG ATCTCAGGCT AGGAGCTCAG GCAACATCTT AGTCCGGTCT CATCTGTTCC TGGATGTCCC TCAGACCCCC
8000  AGCTTTCATC TTTTAGGATT TATTCCTTCC CTGGGATAAT ATAATTTGTG GTCCAAAAAG AACATCATCA AAATTTCAGG CAGAATGGGC CAGGAAGGCC
8100  ATTCTTTCTT GATGAGTGTC CCCAAATCAT CTCCAATTAA CAGACAAGGA GCTTGAGGTT AGGGAGGTGA GGGTAACACT GTCTGTAAGA GGCAGAGCTG
8200  GGACTCAAAT TCCAGATTTC AGATTCCAAA TCCCATCGTT TTTTATCTCT ACAATGATGC CTCCCATCTG GGTGGTGGAG AGAAGGGAGG CGTGTAAAAT
8300  GTCAGCCCCA GAAGGACAAG AGCAAGCCAG TGTGAGCGGA ATTGATGGCT GCAAGCTGAG ACTTGGATTG GAGACGTAGT GAGACTCAGG ATTGTGCAGT
8400  GTCTGCAGGG AAGTGGTTGC TGGATAGAGG CATGGGCTGA ACCAAGCAGC TGGACTGAGA CTGGGGGACA GAACTCCAAA GCCCACTGAG ATGTGGGAAA
8500  ACATGGAGAA GCACACGGAG CATTCACAAC TTATTGCCGT CAGAGTCAAT ACATGGGTGA GGTGGGGATT GGGCAAGAGG GAAAGCGTCA GCCTTCCCTG
8600  ATATTCTGGA AAGTCTCCCG GGGCTGGGGG TGGGCAGGTA CAGAGCTTCG AGCTCTGCTG ATCGCTGACA TCCAGGGGTG GGGGTAGGAA GAGACCTGGG

                                                spProProS erValValVa lThrSerHis GlnAlaProG lyGluLysLy
8700  CCGGGAGAAG TCCACCTCAA GCCTGCAGTG TCACACTCTA TCCCTCCACA GATCCTCCCT CTGTGGTGGT CACCAGCCAC CAGGCCCCAG GAGAAAAGAA

      sLysLeuLys CysLeuAlaT yrAspPheTy rProGlyLys IleAspValH isTrpThrAr gAlaGlyGlu ValGlnGluP roGluLeuAr gGlyAspVal
8800  GAAACTGAAG TGCCTGGCCT ACGACTTCTA CCCAGGGAAA ATTGATGTGC ACTGGACTCG GGCCGGCGAG GTGCAGGAGC CTGAGTTACG GGGAGATGTT

      LeuHisAsnG lyAsnGlyTh rTyrGlnSer TrpValValV alAlaValPr oProGlnAsp ThrAlaProT yrSerCysHi sValGlnHis SerSerLeuA
8900  CTTCACAATG GAAATGGCAC TTACCAGTCC TGGGTGGTGG TGGCAGTGCC CCCGCAGGAC ACAGCCCCCT ACTCCTGCCA CGTGCAGCAC AGCAGCCTGG Exon 4

      laGlnProLe uValValPro TrpGluAlaS erEND
9000  CCCAGCCCCT CGTGGTGCCC TGGGAGGCCA GCTAGGAAGC AAGGGTTGGA GGCAATGTGG GATCTCAGAC CCAGTAGCTG CCCTTCCTGC CTGATGTGGG
9100  AGCTGAACCA CAGAAATCAC AGTCAATGGA TCCACAAGGC CTGAGGAGCA GTGTGGGGGG ACAGACAGGA GGTGGATTTG GAGACCGAAG ACTGGGATGC
9200  CTGTCTTGAG TAGACTTGGA CCCAAAAAAT CATCTCACCT TGAGCCCACC CCCACCCCAT TGTCTAATCT GTAGAAGCTA ATAAATAATC ATCCCTCCTT
        +9306
9300  GCCTAGCATA ACAGAGAATC CTTTTTTTAA CGGTGATGCG CTGTAGAAAT GTGACTAGAT TTTCTCATTG GTTCTGCCCT CAAGCACTGA ATTC
```

FIGURE 3: Nucleotide sequence of the human Zn-α_2-glycoprotein gene. The transcription starting site determined by primer extension (Figure 4) is designated as nucleotide 1. The exons for the protein-coding region are translated into amino acids. Possible transcription-regulating sequences (TATA box, CAT box, an octamer motif, and three Sp1-binding sites) and 13 *Alu* repeats are underlined. Exon 1 of pseudogene 1 is doubly underlined.

Biochemical Corp. (Cleveland, OH). The subcloning was intended for sequencing on both strands everywhere. Sequencing reactions with [α-^{32}P]dCTP were carried out essentially according to the sequencing manual provided by the company. The reaction mixtures were applied usually three times at 2-h intervals to 7% polyacrylamide sequencing gels prepared in 0.5× TBE buffer and run at 40 W for a total of 6 h. The gel was then dried and exposed overnight to X-ray film (RX, Fuji Photo Film Co. Ltd., Minamiashigara, Japan). The DNA sequence data were analyzed with the program GENETYX (Software Development Co. Ltd., Tokyo, Japan) by Dr. H. Mori at the Institute for Virus Research of Kyoto University.

Primer Extension Analysis. Poly(A) RNA was purified with oligo(dT)–cellulose (P-L Pharmacia, Upsala, Sweden) from total RNA of a human liver specimen according to the method in *Molecular Cloning: A laboratory manual* (Sambrook et al., 1989). The nucleotide sequence of the synthesized DNA primer was 5′-CATCTTGGTTCTCCTGGGGG-3′. The DNA (1.84 μg) was labeled with [γ-^{32}P]ATP and

polynucleotide kinase (Toyobo) at 37 °C for 1 h. The labeled primer was precipitated with ethanol three times, dried, and dissolved in 10 μL of distilled water. The human liver poly(A) RNA (10 μg) was annealed with the labeled primer (8 μL) in 10 mM Tris-HCl (pH 8.3) buffer containing 1 mM EDTA and 0.25 M KCl by decreasing the temperature of the mixture from 60 to 42 °C gradually (1.5 h). The primer-extension reaction was carried out in 20 mM Tris-HCl buffer (pH 8.3) containing 75 mM KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, 0.25 mM EDTA, 250 μM each of dNTPs, 100 μg/mL actinomycin D, 200 units of RNasin (Takara), and 20 units of AMV reverse transcriptase (Takara) at 42 °C for 1 h. The products were precipitated with ethanol three times, dried, dissolved in 90% formamide containing 0.5× TBE and dye (bromophenol blue and xylene cyanol), heat-denatured, and loaded onto a sequencing gel. The size marker was the sequence ladder for the subclone R39, which contains the 603-bp *Pvu*II fragment derived from the *Eco*RI c fragment (Figures 1 and 2), when the same primer as above was used.

Southern Blot Analysis of Human DNA. High molecular weight DNA isolated from human peripheral blood leukocytes was digested with *Eco*RI, *Eco*RI + *Bam*HI, *Bam*HI, *Bam*HI + *Pst*I, *Pst*I, or *Pst*I + *Eco*RI (5 μg of each); separated on a 0.7% agarose gel; and blotted to nitrocellulose filter (Southern, 1975). The hybridization probe was the 435-bp *Eco*RI fragment of the human cDNA for Zn-α_2-glycoprotein. The probe corresponds to the region from the 5′-end to the 126th amino acid (Glu) and can detect exons 1, 2, and 3.

Fluorescence in Situ Hybridization. High-resolution R-banded chromosomes were made from human peripheral lymphocytes, which had been cultured in the presence of phytohemagglutinin for 72 h, synchronized with excess thymidine (25 μg/mL for 2 h), and then labeled with BrdU by the method of Viegas-Péquignot and Dutrillaux (1978). Phage DNA (ZG29 or ZG1, 0.5 μg) was labeled by nick translation using biotin-16–dUTP according to the directions of the supplier (Boehringer Mannheim, Germany), hybridized with sonicated total human DNA (50 μg) at 37 °C for 4 h to suppress signals due to repetitive sequences, and then used as a probe. *In situ* hybridization and rinsing was carried out according to the protocol of Takahashi et al. (1990). The hybridized signals amplified with avidin–FITC conjugate (Boehringer Mannheim) were detected under an epifluorescence microscope (Olympus, Tokyo, Japan) with an EO530 filter. Fujichrome film (Fuji; ASA 400) was used for photomicrography.

RESULTS AND DISCUSSION

Screening of Human Genomic Library. A total of 8×10^5 plaques were screened for the human Zn-α_2-glycoprotein gene, and 59 positive clones were obtained. Their DNA was isolated and subjected to *Eco*RI digestion, electrophoresis, and blotting (Southern, 1975). The electrophoretic pattern led to the classification of the phages into six groups, representatives of which, ZG1, ZG2, ZG3, ZG10, ZG12, and ZG29, were picked up. ZG1 and -2 and ZG10 and -12 had overlapping inserts. In order to analyze these phages in detail, three hybridization probes, N, M, and C, were prepared from the 5′-region, the middle region, and the 3′-region respectively, of the cDNA for human Zn-α_2-glycoprotein. ZG1, and ZG2, and ZG29 were hybridized only with N probe; ZG3 and ZG10, only with C probe; and ZG12, with all three probes. More detailed analysis showed that four of them (ZG3, -10, -12, and -29) were overlapped clones. The results are summarized in Figure 1, which shows two genes linked in a head-to-head orientation.

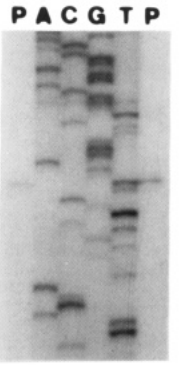


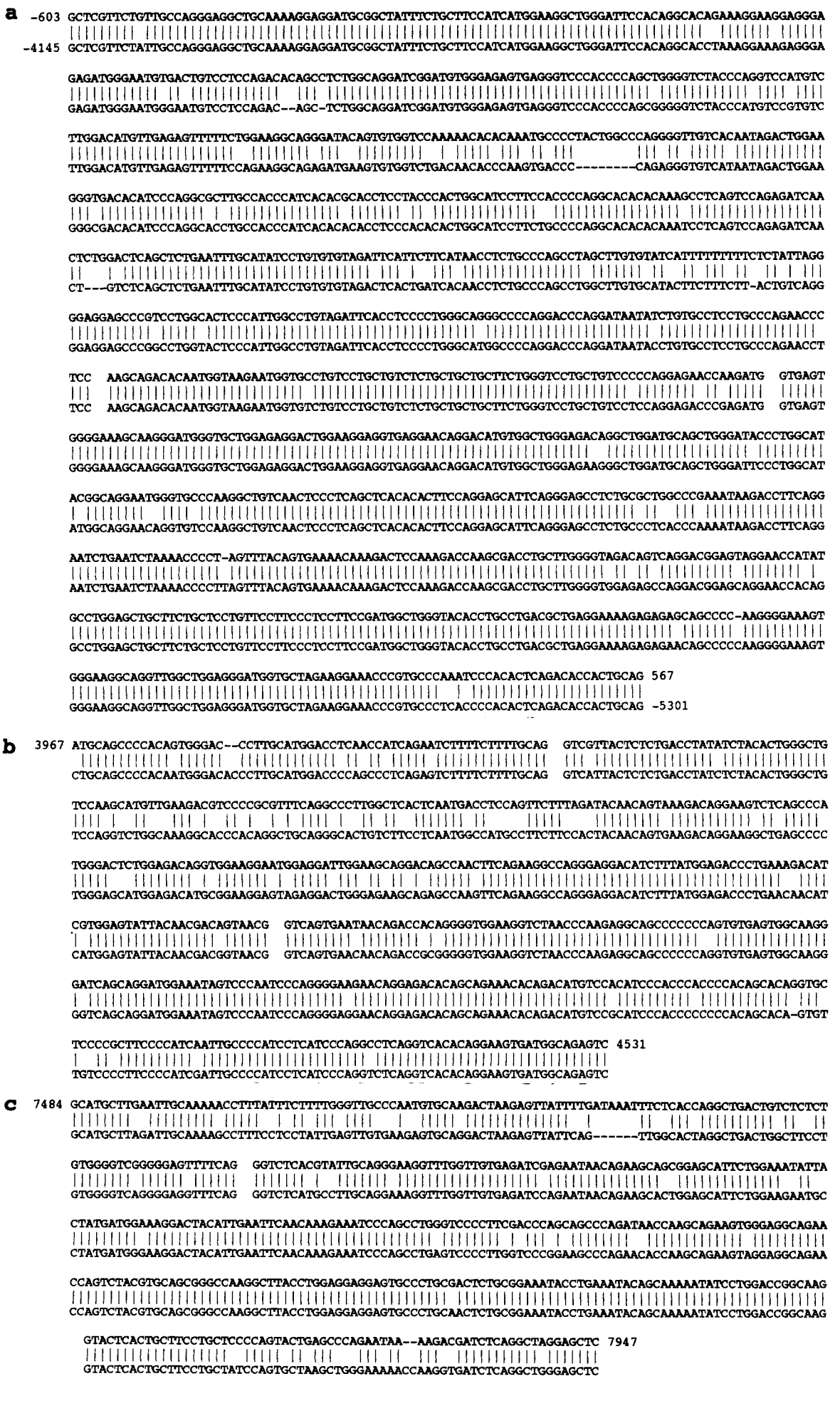
FIGURE 4: Primer extension analysis. Poly(A) RNA isolated from human liver was annealed with ^{32}P-labeled primer DNA (5′-CATCTTGGTTCTCCTGGGGG-3′) and reverse-transcribed. The product was loaded onto a sequencing gel (lane P). A, C, G, and T are the sequence ladder for the subclone R39, the insert of which is the 603-bp *Pvu*II gene fragment covering exon 1, when the same primer as above was used.

The right-hand one is the functional gene, and the left-hand one is a pseudogene (pseudogene 1). ZG1 and -2 were also later found to be clones for a different pseudogene (pseudogene 2); it has a nucleotide deletion in the coding region for the 36th and 37th amino acids resulting in a frameshift. However, we do not refer to this pseudogene here in detail, because its 3rd and 4th exons have not been cloned yet.

Nucleotide Sequence of the Gene for Human Zn-α_2-glycoprotein. The precise structure of the gene for human Zn-α_2-glycoprotein and the sequencing strategy are shown in Figure 2. The sequence analysis was performed 100% on both strands. The entire nucleotide sequence for the functional gene is shown in Figure 3. The results of primer extension analysis (Figure 4) indicated that the transcription starting site is A, 11 nucleotides upstream from the initiation codon ATG, which is designated as nucleotide 1 in Figure 3. The ATG codon is 9 nucleotides upstream from the putative initiation codon in the previous report (Ueyama et al., 1991), indicating that the signal peptide consists of 20 amino acids, not 17. The gene size from the cap site to the poly(A) addition site is 9306 bp. Freije et al. (1991) also reported the nucleotide sequence of cDNA of human Zn-α_2-glycoprotein from breast tissues, but the 5′-region (about 35 nucleotides) of the cDNA is not found in the gene. The discrepancy must result from an artifact in their cDNA cloning.

The gene consists of four exons and three introns. The first exon (exon 1) is for the region from the cap site to the 6th amino acid (Gly); the second exon (exon 2), for the region from the 6th to the 93rd amino acid (Gly); the third exon (exon 3), for the region from the 93rd to the 185th amino acid (Asp); and the fourth exon (exon 4), for the region from the 185th amino acid to the end of the mRNA.

The nucleotide sequence of the exons is identical to the one reported for the cDNA (Ueyama et al., 1991) except for two positions: nucleotide 16 in the signal sequence coding region is T to give a codon for Val (C to give Ala in the cDNA), and the codon for the 84th amino acid is ATC (still for Ile, as in ATT in the cDNA). We must emphasize the fact here that human Zn-α_2-glycoprotein is composed of 278, not 276, amino acids, because we preserved the amino acid numbering proposed by Araki et al. (1988) in the previous report (Ueyama et al., 1991) in order to correct their data; two amino acids (Ile-Phe) must be inserted between the 75th (Asp) and 76th (Met) amino acids. Moreover, the protein in human seminal plasma is composed of 277 amino acids, because it lacks the


a
-603 GCTCGTTCTGTTGCCAGGGAGGCTGCAAAAGGAGGATGCGGCTATTTCTGCTTCCATCATGGAAGGCTGGGATTCCACAGGCACAGAAAGGAAGGAGGGA
-4145 GCTCGTTCTATTGCCAGGGAGGCTGCAAAAGGAGGATGCGGCTATTTCTGCTTCCATCATGGAAGGCTGGGATTCCACAGGCACCTAAAGGAAAGAGGGA
GAGATGGGAATGTGACTGTCCTCCAGACACAGCCTCTGGCAGGATCGGATGTGGGAGAGTGAGGGTCCCACCCCAGCTGGGGTCTACCCAGGTCCATGTC
GAGATGGGAATGGGAATGTCCTCCAGAC--AGC-TCTGGCAGGATCGGATGTGGGAGAGTGAGGGTCCCACCCCAGCGGGGGTCTACCCATGTCCGTGTC
TTGGACATGTTGAGAGTTTTTCTGGAAGGCAGGGATACAGTGTGGTCCAAAAACACACAAATGCCCCTACTGGCCCAGGGGTTGTCACAATAGACTGGAA
TTGGACATGTTGAGAGTTTTTCCAGAAGGCAGAGATGAAGTGTGGTCTGACAACACCCAAGTGACCC--------CAGAGGGTGTCATAATAGACTGGAA
GGGTGACACATCCCAGGCGCTTGCCACCCATCACACGCACCTCCTACCCACTGGCATCCTTCCACCCCAGGCACACACAAAGCCTCAGTCCAGAGATCAA
GGGCGACACATCCCAGGCACCTGCCACCCATCACACACACCTCCCACACACTGGCATCCTTCTGCCCCAGGCACACACAAATCCTCAGTCCAGAGATCAA
CTCTGGACTCAGCTCTGAATTTGCATATCCTGTGTGTAGATTCATTCTTCATAACCTCTGCCCAGCCTAGCTTGTGTATCATTTTTTTTTCTCTATTAGG
CT---GTCTCAGCTCTGAATTTGCATATCCTGTGTGTAGACTCACTGATCACAACCTCTGCCCAGCCTGGCTTGTGCATACTTCTTTCTT-ACTGTCAGG
GGAGGAGCCCGTCCTGGCACTCCCATTGGCCTGTAGATTCACCTCCCCTGGGCAGGGCCCCAGGACCCAGGATAATATCTGTGCCTCCTGCCCAGAACCC
GGAGGAGCCCGGCCTGGTACTCCCATTGGCCTGTAGATTCACCTCCCCTGGGCATGGCCCCAGGACCCAGGATAATACCTGTGCCTCCTGCCCAGAACCT
TCC AAGCAGACACAATGGTAAGAATGGTGCCTGTCCTGCTGTCTCTGCTGCTGCTTCTGGGTCCTGCTGTCCCCCAGGAGAACCAAGATG GTGAGT
TCC AAGCAGACACAATGGTAAGAATGGTGTCTGTCCTGCTGTCTCTGCTGCTGCTTCTGGGTCCTGCTGTCCTCCAGGAGACCCGAGATG GTGAGT
GGGGAAAGCAAGGGATGGGTGCTGGAGAGGACTGGAAGGAGGTGAGGAACAGGACATGTGGCTGGGAGACAGGCTGGATGCAGCTGGGATACCCTGGCAT
GGGGAAAGCAAGGGATGGGTGCTGGAGAGGACTGGAAGGAGGTGAGGAACAGGACATGTGGCTGGGAGAAGGGCTGGATGCAGCTGGGATTCCCTGGCAT
ACGGCAGGAATGGGTGCCCAAGGCTGTCAACTCCCTCAGCTCACACACTTCCAGGAGCATTCAGGGAGCCTCTGCGCTGGCCCGAAATAAGACCTTCAGG
ATGGCAGGAACAGGTGTCCAAGGCTGTCAACTCCCTCAGCTCACACACTTCCAGGAGCATTCAGGGAGCCTCTGCCCTCACCCAAAATAAGACCTTCAGG
AATCTGAATCTAAAACCCCT-AGTTTACAGTGAAAACAAAGACTCCAAAGACCAAGCGACCTGCTTGGGGTAGACAGTCAGGACGGAGTAGGAACCATAT
AATCTGAATCTAAAACCCCTTAGTTTACAGTGAAAACAAAGACTCCAAAGACCAAGCGACCTGCTTGGGGTGGAGAGCCAGGACGGAGCAGGAACCACAG
GCCTGGAGCTGCTTCTGCTCCTGTTCCTTCCCTCCTTCCGATGGCTGGGTACACCTGCCTGACGCTGAGGAAAAGAGAGAGCAGCCCC-AAGGGGAAAGT
GCCTGGAGCTGCTTCTGCTCCTGTTCCTTCCCTCCTTCCGATGGCTGGGTACACCTGCCTGACGCTGAGGAAAAGAGAGAACAGCCCCCAAGGGGAAAGT
GGGAAGGCAGGTTGGCTGGAGGGATGGTGCTAGAAGGAAACCCGTGCCCAAATCCCACACTCAGACACCACTGCAG 567
GGGAAGGCAGGTTGGCTGGAGGGATGGTGCTAGAAGGAAACCCGTGCCCTCACCCCACACTCAGACACCACTGCAG -5301
b
3967 ATGCAGCCCCACAGTGGGAC--CCTTGCATGGACCTCAACCATCAGAATCTTTTCTTTTGCAG GTCGTTACTCTCTGACCTATATCTACACTGGGCTG
CTGCAGCCCCACAATGGGACACCCTTGCATGGACCCCAGCCCTCAGAGTCTTTTCTTTTGCAG GTCATTACTCTCTGACCTATCTCTACACTGGGCTG
TCCAAGCATGTTGAAGACGTCCCCGCGTTTCAGGCCCTTGGCTCACTCAATGACCTCCAGTTCTTTAGATACAACAGTAAAGACAGGAAGTCTCAGCCCA
TCCAGGTCTGGCAAAGGCACCCACAGGCTGCAGGGCACTGTCTTCCTCAATGGCCATGCCTTCTTCCACTACAACAGTGAAGACAGGAAGGCTGAGCCCC
TGGGACTCTGGAGACAGGTGGAAGGAATGGAGGATTGGAAGCAGGACAGCCAACTTCAGAAGGCCAGGGAGGACATCTTTATGGAGACCCTGAAAGACAT
TGGGAGCATGGAGACATGCGGAAGGAGTAGAGGACTGGGAGAAGCAGAGCCAAGTTCAGAAGGCCAGGGAGGACATCTTTATGGAGACCCTGAACAACAT
CGTGGAGTATTACAACGACAGTAACG GTCAGTGAATAACAGACCACAGGGGTGGAAGGTCTAACCCAAGAGGCAGCCCCCCCAGTGTGAGTGGCAAGG
CATGGAGTATTACAACGACGGTAACG GTCAGTGAACAACAGACCGCGGGGGTGGAAGGTCTAACCCAAGAGGCAGCCCCCCAGGTGTGAGTGGCAAGG
GATCAGCAGGATGGAAATAGTCCCAATCCCAGGGGAAGAACAGGAGACACAGCAGAAACACAGACATGTCCACATCCCACCCACCCCACAGCACAGGTGC
GGTCAGCAGGATGGAAATAGTCCCAATCCCAGGGGAGGAACAGGAGACACAGCAGAAACACAGACATGTCCGCATCCCACCCCCCCCACAGCACA-GTGT
TCCCCGCTTCCCCATCAATTGCCCCATCCTCATCCCAGGCCTCAGGTCACACAGGAAGTGATGGCAGAGTC 4531
TGTCCCCTTCCCCATCGATTGCCCCATCCTCATCCCAGGTCTCAGGTCACACAGGAAGTGATGGCAGAGTC
c
7484 GCATGCTTGAATTGCAAAAACCTTTATTTCTTTTGGGTTGCCCAATGTGCAAGACTAAGAGTTATTTTGATAAATTTCTCACCAGGCTGACTGTCTCTCT
GCATGCTTAGATTGCAAAAGCCTTTCCTCCTATTGAGTTGTGAAGAGTGCAGGACTAAGAGTTATTCAG------TTGGCACTAGGCTGACTGGCTTCCT
GTGGGGTCGGGGGAGTTTTCAG GGTCTCACGTATTGCAGGGAAGGTTTGGTTGTGAGATCGAGAATAACAGAAGCAGCGGAGCATTCTGGAAATATTA
GTGGGGTCAGGGGAGGTTTCAG GGTCTCATGCCTTGCAGGAAAGGTTTGGTTGTGAGATCCAGAATAACAGAAGCACTGGAGCATTCTGGAAGAATGC
CTATGATGGAAAGGACTACATTGAATTCAACAAAGAAATCCCAGCCTGGGTCCCCTTCGACCCAGCAGCCCAGATAACCAAGCAGAAGTGGGAGGCAGAA
CTATGATGGGAAGGACTACATTGAATTCAACAAAGAAATCCCAGCCTGAGTCCCCTTGGTCCCGGAAGCCCAGAACACCAAGCAGAAGTAGGAGGCAGAA
CCAGTCTACGTGCAGCGGGCCAAGGCTTACCTGGAGGAGGAGTGCCCTGCGACTCTGCGGAAATACCTGAAATACAGCAAAAATATCCTGGACCGGCAAG
CCAGTCTACGTGCAGCGGGCCAAGGCTTACCTGGAGGAGGAGTGCCCTGCAACTCTGCGGAAATACCTGAAATACAGCAAAAATATCCTGGACCGGCAAG
GTACTCACTGCTTCCTGCTCCCCAGTACTGAGCCCAGAATAA--AAGACGATCTCAGGCTAGGAGCTC 7947
GTACTCACTGCTTCCTGCTATCCAGTGCTAAGCTGGGAAAAACCAAGGTGATCTCAGGCTGGGAGCTC

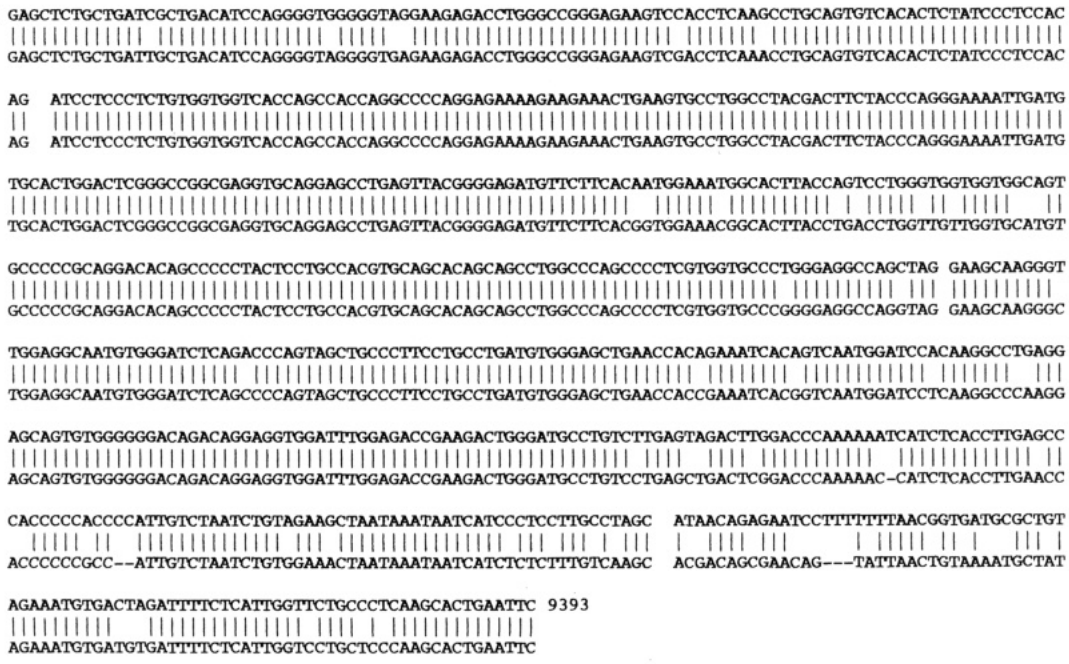


FIGURE 5: Comparison of nucleotide sequence between the functional gene for Zn-α_2-glycoprotein and pseudogene 1. Identical bases are indicated with vertical bars; the upper rows are the functional gene, and the lower rows are pseudogene 1. The numbers correspond to those in Figure 3. The regions between two-base blanks are the exons; sequences a, b, c, and d are around exons 1, 2, 3, and 4, respectively. From the one-base blank downstream to the two-base blank in d is the 3′-untranslated region.

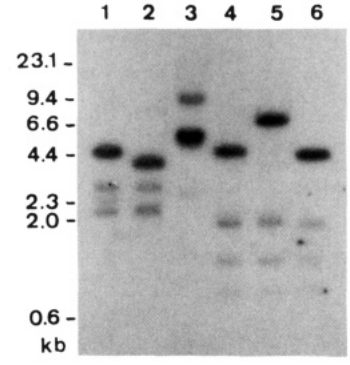


FIGURE 6: Southern blot of human genomic DNA. High molecular weight human DNA was digested with *Eco*RI (1), *Eco*RI + *Bam*HI (2), *Bam*HI (3), *Bam*HI + *Pst*I (4), *Pst*I (5), or *Pst*I + *Eco*RI (6); separated by electrophoresis; and blotted onto nitrocellulose filter. The hybridization probe was the 435-bp *Eco*RI fragment derived from the cDNA for human Zn-α_2-glycoprotein. The probe corresponds to the region from the 5′-end to the 126th amino acid and can detect exons 1, 2, and 3. The standard is λ phage DNA cut with *Hin*dIII.

Table I: DNA Fragments Detectable by Southern Blot[a]

	restriction endonucleases					
genes	*Eco*RI	*Eco*RI + *Bam*HI	*Bam*HI	*Bam*HI + *Pst*I	*Pst*I	*Pst*I + *Eco*RI
functional						
exon 1	3.1	3.1	>10	2.0	2.0	2.0
exons 2 and 3	4.6	4.1	5.5	4.8	7.8	4.6
pseudogene 1						
exon 1	2.2	2.2	>10	1.1	1.1	1.1
exon 2	2.4	2.2	5.0	0.8	0.8	0.8
				0.1	0.1	0.1
exon 3	0.7	0.7	5.0	2.0	2.0	0.7
pseudogene 2						
exon 1	2.2	2.2	2.7	1.1	1.1	1.1
exon 2	>4.0	>3.8	>3.8	1.1	1.1	1.1
				0.2	0.5	0.5

[a] The hybridization probe was the 435-bp *Eco*RI fragment derived from the cDNA for Zn-α_2-glycoprotein, which can detect exons 1, 2, and 3. The sizes of the DNA fragments are in kb. The third exon of pseudogene 2 is not cloned yet.

N-terminal amino acid (Gln or pyroglutamic acid) (Ohkubo et al., 1990).

Zn-α_2-glycoprotein was reported to resemble MHC class I α chains in terms of amino acid sequence and domain structure (Araki et al., 1988). Zn-α_2-glycoprotein has 36–39% homology in amino acid sequence with MHC class I α chains (Araki et al., 1988). The N-terminal domain of MHC class I α chains (α1, up to the 90th amino acid) and that of Zn-α_2-glycoprotein (domain A, from the 6th to the 92nd amino acid) are both devoid of cysteine residues. In contrast, the other domains of MHC class I α chains, α2 (from the 91st to the 182nd amino acid) and α3 (from the 183rd to the 274th amino acid), and of Zn-α_2-glycoprotein, B (from the 93rd to the 185th amino acid) and C (from the 186th to the 278th amino acid), have two cysteine residues each to make disulfide bonds within themselves. In the case of MHC class I α chains, each of these three extracellular domains, α1, α2, and α3, is precisely encoded by a separate exon (Malissen et al., 1982). Zn-α_2-glycoprotein is similar to MHC in this respect; each of the three domains, A, B and C, is encoded by a separate exon. However, exon 4 of MHC Class I α chain genes corresponds to domain α3 only, while exon 4 of the Zn-α_2-glycoprotein gene corresponds to both domain C and the 3′-untranslated region. In the MHC class I α chain genes, the 3′-untranslated region is encoded by exon 8 together with the cytoplasmic domain 3. It was plausible that domain C and the 3′-untranslated region are encoded by a separate exon in the Zn-α_2-glycoprotein gene, but this is not the case. Zn-α_2-glycoprotein is demonstrated here not to be an alternatively processed product of a gene whose exon–intron structure is the same as that of MHC class I α chain genes. However, this does not exclude the membership of Zn-α_2-glycoprotein in the immunoglobulin gene superfamily (Hunkapiller & Hood, 1986). At the nucleotide level, the protein-coding region of the Zn-α_2-glycoprotein gene (the region from the 6th to the 278th amino acid) has 52.8% homology with the consensus nucleotide sequence of cDNA (from the 1st (Gly) to the 278th (Ser) amino acid) proposed for HLA-A and -B by Ennis et

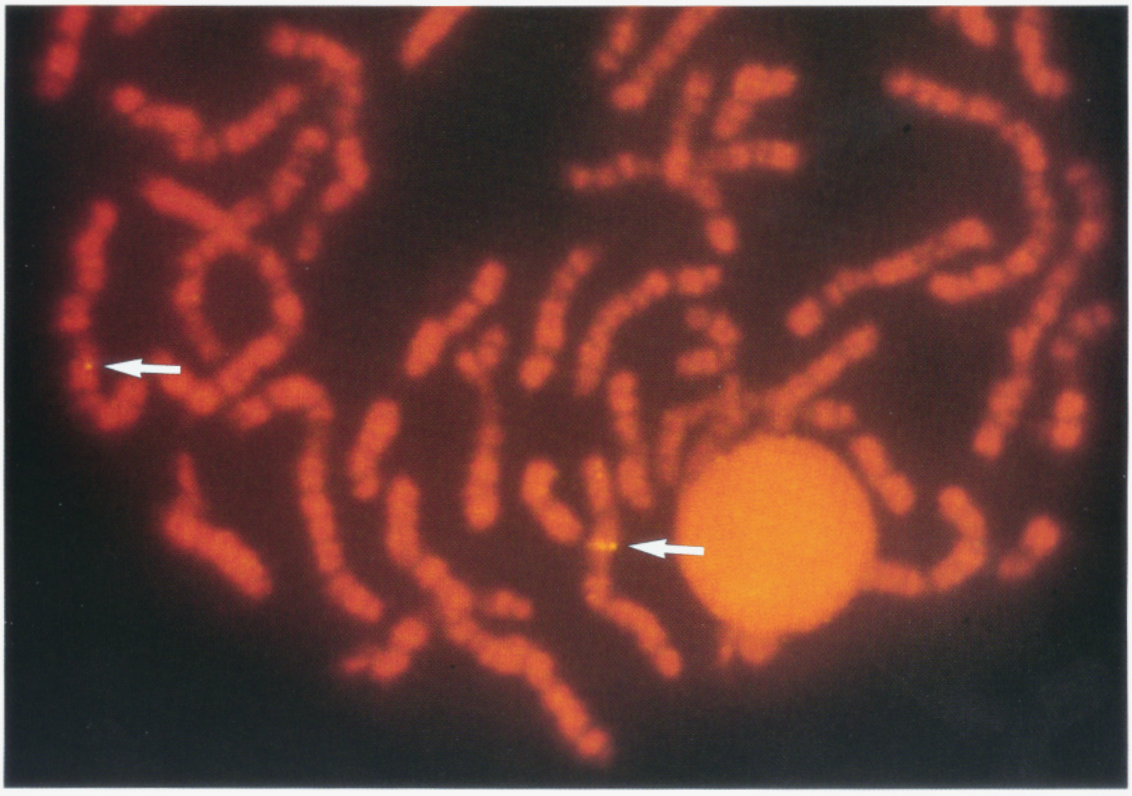


FIGURE 7: Chromosomal localization of the ZA2G gene by fluorescence *in situ* hybridization. Arrows indicate fluorescent signals for the ZA2G gene on both chromatids at band 7q22.1.

al. (1990). There is no significant homology between these two genes in the first exon or in the 3′-untranslated region.

Ten copies of *Alu* repetitive DNA were identified in the introns, and these comprise about 30% of the gene. There are three more copies of *Alu* sequence in the 5′-flanking region, but they are all incomplete. Including these *Alu* repeats, we designated the most upstream one as *Alu* 1. It has been reported that *Alu* repeats are dispersed throughout the human genome with over 500 000 copies (Schmid & Jelinek, 1982), indicating that the repeat is found every 3–5 kb of the genome. Therefore, the presence of 10 repeats of *Alu* DNA within the human Zn-α_2-glycoprotein gene of 9.3-kb size is considered to be frequent. Such a high content of *Alu* repeats has been found in few genes, such as the genes for human prothrombin (40%; Degen & Davie, 1987) and glucose-6-phosphate dehydrogenase (24%; Chen et al., 1991). All of the *Alu* repeats inside the gene belong to the S subfamily except for *Alu* 13, which belongs to the J subfamily, the oldest class of *Alu* repeat (Jurka & Smith, 1988; Britten et al., 1988). Two monomer *Alu* repeats, *Alu* 10 (right monomer) and *Alu* 12 (left monomer), are considered to have been originally a complete dimer but to have become interrupted by *Alu* 11 in the linker sequence (TAATTTT) sometime during the evolution. The *Alu* 10–*Alu* 13 cluster is flanked by direct repeats, but *Alu* 10 and *Alu* 13 themselves are not flanked by such repeats, as follows: TGTAGCT – T tail–*Alu* 10–linker–T tail–*Alu* 11–linker–*Alu* 12–T tail–*Alu* 13 – TGTAGCT. If the presence of *Alu* repeats in tandem is simply due to independent insertion of different *Alu* copies as suggested by Chen et al. (1991), the most plausible order of their insertion into the gene is first *Alu* 13, then *Alu* 10 and *Alu* 12 as a complete dimer into the T tail, and finally *Alu* 11 into the linker between *Alu* 10 and *Alu* 12. In addition to this insertion model, we must consider another mechanism, i.e., duplication or unequal crossover of an inserted *Alu* repeat. Since *Alu* 13 belongs to a different subfamily (J) from *Alu* repeats 10–12 (all S), the presence of *Alu* repeats in tandem is not ascribed solely to the second mechanism. Studies of corresponding genes in other primates should be helpful to know how the *Alu* cluster arose.

In the 5′-flanking region of the gene, TATA and CAT boxes are found; the TATA box (TAATAT) is 26 nucleotides upstream from the cap site, and the CAT box in inverted orientation (ATTGG) is 75 nucleotides upstream from the cap site (Figure 3). In addition to these, an octamer sequence (ATTTGCAT, at nucleotide –178) and three possible Sp1-binding sites (at nucleotides –47, –97, and –130), which may be needed for a basal level of transcription of this protein, were found. Since the 5′-flanking region is abruptly interrupted by an incomplete *Alu* repeat, *Alu* 3, unequal crossover of chromosome 7 around the gene (Ueyama et al., 1991) may have involved this *Alu* sequence, which is not present in the 5′-region of pseudogene 1. Hence, regulatory elements for the transcription of this gene are considered to reside in the region between the terminus of *Alu* 3 and the cap site (601 nucleotides).

Comparison of Nucleotide Sequence between the Functional Gene and Pseudogene 1. The nucleotide sequence of pseudogene 1 contained in ZG10 and ZG12 was also determined for the exons and their adjacent introns. We concluded that the gene is nonfunctional, because the codons for the 134th and 148th amino acids are both TAG (stop codon), not TGG for Trp as in the functional counterpart. However, the difference in nucleotide sequence between the two genes is very small (Figure 5). It is 10% (88/885) in the amino acid coding region, 9% (25/275) in the 3′-untranslated region, and also 9% (96/1080) in the introns adjacent to the exons so far sequenced. The same degree of difference among these regions suggests that the divergence began immediately after the duplication of the gene. There is no insertion or deletion of nucleotides to cause a frameshift in the protein-coding region.

The poly(A) addition (AATAAA) and splicing (GT–AG) signals are also conserved in the pseudogene. These observations suggest that the duplication of the gene occurred quite recently in evolution. If we employ the rate of evolution estimated by Hayashida and Miyata (1983) (5.3×10^{-9}/site/year), the 9% difference (9.6% corrected difference for multiple substitutions) means the divergence started 18 million years ago. Since the divergence date of the orangutan lineage and the lineage to man is believed to be 13 million years ago (Raza et al., 1983), comparison of this type of pseudogenes for Zn-α_2-glycoprotein among primates may be of interest.

Southern Blot of Human DNA. In the previous report (Ueyama et al., 1991), we suggested from a simple pattern of the genomic Southern blot that the gene for Zn-α_2-glycoprotein is a single copy in the human genome, but we have already found two pseudogenes. Southern blot analysis was therefore performed again with another probe. This time we used the 435-bp *Eco*RI fragment of the cDNA, which corresponds to exons 1, 2, and 3 and was expected to detect the fragments listed in Table I. The results shown in Figure 6 are consistent with the prediction, suggesting that in humans there are only three genes for Zn-α_2-glycoprotein, which we have cloned.

Chromosomal Mapping of the Human Zn-α_2-glycoprotein Gene (ZA2G). Using a panel of rodent–human somatic cell hybrids, the Zn-α_2-glycoprotein gene was assigned to human chromosome 7 (Ueyama et al., 1991). In this study regional assignment of the gene was carried out by fluorescence *in situ* hybridization. Among the typical R-banded chromosomes we screened (50 metaphase preparations), 38% had fluorescent signals symmetrically on both chromatids at 7q22.1 (Figure 7), and 42% had one signal at the same band on either chromatid. Thus, the ZA2G gene was assigned to the 7q22.1 band of the human karyotype. The same results were obtained when ZG1 phage DNA was used as a probe, suggesting that pseudogene 2 is also closely linked to the functional gene on the chromosome.

We report here the entire nucleotide sequence of the gene for human Zn-α_2-glycoprotein. The physiological role(s) of this protein is (are) not known yet, but its resemblance to MHC antigens (Araki et al., 1988) and its possible function as a carrier protein of nephritogenic glycoprotein (Shibata & Miura, 1982) suggest a role in the immune response. The nucleotide sequence data we present here will be useful when disorders closely related to abnormalities of this protein are found and for analyses of the control mechanisms of exocrine gland specific transcription and also of the microheterogeneity of this protein (Nakayashiki & Katsura, 1989; Ding et al., 1990).

ACKNOWLEDGMENT

We are grateful to Dr. Masaki Takiguchi for his providing us with a human genomic library. We also thank Dr. Hirotada Mori for computer analysis of the nucleotide sequence. Further, we thank Ms. Rie Ohsugi for her technical assistance.

REFERENCES

Araki, T., Gejyo, F., Takagaki, K., Haupt, H., Schwick, H. G., Bürgi, W., Marti, T., Schaller, J., Rickli, E., Brossmer, R., Atkinson, P. H., Putnam, F. W., & Schmid, K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 679–683.

Benton, W. D., & Davis, R. W. (1977) *Science 196*, 180–182.

Britten, R. J., Baron, W. F., Stout, D. B., & Davidson, E. H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4770–4774.

Bürgi, W., & Schmid, K. (1961) *J. Biol. Chem. 236*, 1066–1074.

Chen, E. Y., Cheng, A., Lee, A., Kuang, W.-J., Hillier, L., Green, P., Schlessinger, D., Ciccodicola, A., & D'Urso, M. (1991) *Genomics 10*, 792–800.

Degen, S. J. F., & Davie, E. W. (1987) *Biochemistry 26*, 6165–6177.

Ding, M., Umetsu, K., Nakayashiki, N., Choi, W. Y., Jia, J., & Suzuki, T. (1990) *Hum. Hered. 40*, 311–312.

Ennis, P. D., Zemmour, J., Salter, R. D., & Parham, P. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2833–2837.

Freije, J. P., Fueyo, A., Uría, J., & López-Otín, C. (1991) *Eur. J. Biochem. 290*, 247–249.

Frenette, G., Dubé, J. Y., Lazure, C., Paradis, G., Chrétien, M., & Tremblay, R. R. (1987) *Prostate 11*, 257–270.

Hayashida, H., & Miyata, T. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2671–2675.

Hunkapiller, T., & Hood, L. (1986) *Nature 323*, 15–16.

Jurka, J., & Smith, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4775–4778.

Malissen, M., Malissen, B., & Jordan, B. R. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 893–897.

Nakayashiki, N., & Katsura, S. (1989) *Hum. Genet. 82*, 293–295.

Ohkubo, I., Niwa, M., & Sasaki, M. (1988) *Prep. Biochem. 18*, 413–430.

Ohkubo, I., Niwa, M., Takashima, A., Nishikimi, N., Gasa, S., Sasaki, M. (1990) *Biochim. Biophys. Acta 1034*, 152–156.

Poortmans, J. R., & Schmid, K. (1968) *J. Lab. Clin. Med. 71*, 807–811.

Raza, S. M., Barry, J. C., Pilbeam, D., Rose, M. D., Shah, S. M. I., & Ward, S. (1983) *Nature 306*, 52–54.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A laboratory manual*, Chapter 7, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schmid, C. W., & Jelinek, W. R. (1982) *Science 216*, 1065–1070.

Shibata, S., & Miura, K. (1982) *Nephron 31*, 170–176.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Tada, T., Ohkubo, I., Niwa, M., Sasaki, M., Tateyama, H., & Eimoto, T. (1991) *J. Histochem. Cytochem. 39*, 1221–1226.

Takahashi, E., Hori, T., O'Connell, P., Leppert, M., & White, R. (1990) *Hum. Genet. 86*, 14–16.

Takiguchi, M., Haraguchi, Y., & Mori, M. (1988) *Nucleic Acids Res. 16*, 8789–8802.

Ueyama, H., Niwa, M., Tada, T., Sasaki, M., & Ohkubo, I. (1991) *Biochem. Biophys. Res. Commun. 177*, 696–703.

Viegas-Péquignot, E., & Dutrillaux, B. (1978) *Ann. Genet. 21*, 122–124.